

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/537, 201A
Source: PCT
Date Processed by STIC: 03/06/2006

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:31

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw

```

3 <110> APPLICANT: Van der Geize, Robert
4     Hessels, Gerda
5     Dijkhuizen, Lubbert
6     Van der Meijden, Peter
8 <120> TITLE OF INVENTION: New expression system from Rhodococcus
10 <130> FILE REFERENCE: 2002.744US
12 <140> CURRENT APPLICATION NUMBER: 10/537,201A
13 <141> CURRENT FILING DATE: 2005-06-02
15 <150> PRIOR APPLICATION NUMBER: PCT/EP03/050928
16 <151> PRIOR FILING DATE: 2003-12-02
18 <150> PRIOR APPLICATION NUMBER: EP02080054.6
19 <151> PRIOR FILING DATE: 2002-12-03
21 <160> NUMBER OF SEQ ID NOS: 13
23 <170> SOFTWARE: PatentIn Ver. 2.1
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 1543
27 <212> TYPE: DNA
28 <213> ORGANISM: Rhodococcus erythropolis
30 <220> FEATURE:
31 <221> NAME/KEY: CDS
32 <222> LOCATION: (1)..(1533)
34 <400> SEQUENCE: 1
35 atg cag gac tgg acc agc gag tgc gac gtg ttg gta gtc ggc tcc ggc   48
36 Met Gln Asp Trp Thr Ser Glu Cys Asp Val Leu Val Val Gly Ser Gly
37 1      5      10      15
39 ggc gga gcg ctg acc ggc gca tat acc gcc gct gct cag gga ttg acg   96
40 Gly Gly Ala Leu Thr Gly Ala Tyr Thr Ala Ala Ala Gln Gly Leu Thr
41      20      25      30
43 acg atc gtc ctc gag aaa acc gat cgt ttc ggc ggg acc tcc gcc tac   144
44 Thr Ile Val Leu Glu Lys Thr Asp Arg Phe Gly Gly Thr Ser Ala Tyr
45      35      40      45
47 tcg ggc gcc tcg atc tgg ctc cca ggt acc cag gtg cag gaa cgc gcc   192
48 Ser Gly Ala Ser Ile Trp Leu Pro Gly Thr Gln Val Gln Glu Arg Ala
49      50      55      60
51 gga ctt ccc gac tcg acc gag aat gcc cgc acc tat ctg cgc gcg ttg   240
52 Gly Leu Pro Asp Ser Thr Glu Asn Ala Arg Thr Tyr Leu Arg Ala Leu
53      65      70      75      80
55 ctc ggt gac gcc gag tcc gag cgc cag gac gcc tac gtc gag acc gct   288
56 Leu Gly Asp Ala Glu Ser Glu Arg Gln Asp Ala Tyr Val Glu Thr Ala
57      85      90      95
59 ccc gct gtc gtc gct cta ctc gag cag aac ccg aac atc gaa ttc gag   336
60 Pro Ala Val Val Ala Leu Leu Glu Gln Asn Pro Asn Ile Glu Phe Glu
61      100      105      110

```

RAW SEQUENCE LISTING

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:31

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw

```

63 ttc cgt gcg ttc ccc gac tac tac aaa gcc gaa ggc cgg atg gac acg 384
64 Phe Arg Ala Phe Pro Asp Tyr Tyr Lys Ala Glu Gly Arg Met Asp Thr
65 115 120 125
67 gga cgc tcc atc aac cct ctc gat ctc gat ccc gcc gac atc ggt gac 432
68 Gly Arg Ser Ile Asn Pro Leu Asp Leu Asp Pro Ala Asp Ile Gly Asp
69 130 135 140
71 ctc gcc ggc aag gtg cgt ccg gaa ctg gac caa gac cgc acc ggt cag 480
72 Leu Ala Gly Lys Val Arg Pro Glu Leu Asp Gln Asp Arg Thr Gly Gln
73 145 150 155 160
75 gat cat gct ccc ggc ccg atg atc ggt ggg cgc gca ctg atc ggc cgt 528
76 Asp His Ala Pro Gly Pro Met Ile Gly Gly Arg Ala Leu Ile Gly Arg
77 165 170 175
79 ctg ctg gcc gca gtt cag agc acc ggt aag gca gaa ctt cgc acc gaa 576
80 Leu Leu Ala Ala Val Gln Ser Thr Gly Lys Ala Glu Leu Arg Thr Glu
81 180 185 190
83 tcc gtc ctc acc tcc ctg atc gtg gaa gac ggc cgt gtt gtc ggc gcc 624
84 Ser Val Leu Thr Ser Leu Ile Val Glu Asp Gly Arg Val Val Gly Ala
85 195 200 205
87 gag gtc gaa tcc ggc ggc gaa acc cag cga atc aag gcg aac cgc ggt 672
88 Glu Val Glu Ser Gly Gly Glu Thr Gln Arg Ile Lys Ala Asn Arg Gly
89 210 215 220
91 gtc ctg atg gca gca ggc ggc atc gaa ggc aac gcc gag atg cgt gag 720
92 Val Leu Met Ala Ala Gly Gly Ile Glu Gly Asn Ala Glu Met Arg Glu
93 225 230 235 240
95 cag gca ggc acc ccc ggc aag gcg atc tgg agt atg ggt ccc ttc ggc 768
96 Gln Ala Gly Thr Pro Gly Lys Ala Ile Trp Ser Met Gly Pro Phe Gly
97 245 250 255
99 gcc aac acc ggc gac gcg atc tct gcc ggt att gct gtc ggc ggc gca 816
100 Ala Asn Thr Gly Asp Ala Ile Ser Ala Gly Ile Ala Val Gly Gly Ala
101 260 265 270
103 aca gcc ttg ctc gat cag gcg tgg ttc tgc ccc ggc gtc gag cag ccc 864
104 Thr Ala Leu Leu Asp Gln Ala Trp Phe Cys Pro Gly Val Glu Gln Pro
105 275 280 285
107 gag ggc agc gcc gcc ttc atg gtc ggc gtt cgc ggt ggg ctc gtc gtc 912
108 Asp Gly Ser Ala Ala Phe Met Val Gly Val Arg Gly Gly Leu Val Val
109 290 295 300
111 gac agc gcc ggt gag cgc tac ctc aac gag tcg ctt ccg tac gac cag 960
112 Asp Ser Ala Gly Glu Arg Tyr Leu Asn Glu Ser Leu Pro Tyr Asp Gln
113 305 310 315 320
115 ttc gga cga gcc atg gat gct cac gac gac aac ggt tct gcc gtg ccg 1008
116 Phe Gly Arg Ala Met Asp Ala His Asp Asp Asn Gly Ser Ala Val Pro
117 325 330 335
119 tcg ttc atg atc ttc gac tcg cgc gag ggt ggc gga ctg ccc gcc atc 1056
120 Ser Phe Met Ile Phe Asp Ser Arg Glu Gly Gly Gly Leu Pro Ala Ile
121 340 345 350
123 tgc atc ccg aac acg gcg ccc gcc aag cac ctc gaa gcc gga acg tgg 1104
124 Cys Ile Pro Asn Thr Ala Pro Ala Lys His Leu Glu Ala Gly Thr Trp
125 355 360 365
127 gtc ggt gcc gac act ctc gaa gaa ctc gct gcc aag acc gga cta ccg 1152

```

RAW SEQUENCE LISTING

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:31

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw

```

128 Val Gly Ala Asp Thr Leu Glu Glu Leu Ala Ala Lys Thr Gly Leu Pro
129      370      375      380
131 gcc gac gca ttg cgc agc act gtc gaa aag ttc aac gat gcc gca aaa 1200
132 Ala Asp Ala Leu Arg Ser Thr Val Glu Lys Phe Asn Asp Ala Ala Lys
133 385      390      395      400
135 ctg ggc gtc gac gaa gag ttc cat cgc ggc gaa gac ccg tac gac gcg 1248
136 Leu Gly Val Asp Glu Glu Phe His Arg Gly Glu Asp Pro Tyr Asp Ala
137      405      410      415
139 ttc ttc tgc cca ccc aac ggc ggt gcg aat gcg gca ctg acg gcc atc 1296
140 Phe Phe Cys Pro Pro Asn Gly Gly Ala Asn Ala Ala Leu Thr Ala Ile
141      420      425      430
143 gag aac gga ccg ttc tac gcg gcc cgc atc gtc ctc agt gac ctc ggc 1344
144 Glu Asn Gly Pro Phe Tyr Ala Ala Arg Ile Val Leu Ser Asp Leu Gly
145      435      440      445
147 acc aag ggc gga ttg gtc acc gac gtc aac ggc cga gtc ctg cgt gct 1392
148 Thr Lys Gly Gly Leu Val Thr Asp Val Asn Gly Arg Val Leu Arg Ala
149      450      455      460
151 gac ggc agc gcc atc gac ggc ctg tac gcc gcc ggc aac acg agc gcg 1440
152 Asp Gly Ser Ala Ile Asp Gly Leu Tyr Ala Ala Gly Asn Thr Ser Ala
153 465      470      475      480
155 tca ctg agc ggc cgc ttc tac ccc ggc ccc gga gtt cca ctc ggc acg 1488
156 Ser Leu Ser Gly Arg Phe Tyr Pro Gly Pro Gly Val Pro Leu Gly Thr
157      485      490      495
159 gct atg gtc ttc tcg tac cga gca gct cag gac atg gcg aag taa 1533
160 Ala Met Val Phe Ser Tyr Arg Ala Ala Gln Asp Met Ala Lys
161      500      505      510
163 cgcaattcaa 1543
166 <210> SEQ ID NO: 2
167 <211> LENGTH: 510
168 <212> TYPE: PRT
169 <213> ORGANISM: Rhodococcus erythropolis
171 <400> SEQUENCE: 2
172 Met Gln Asp Trp Thr Ser Glu Cys Asp Val Leu Val Val Gly Ser Gly
173 1 5 10 15
175 Gly Gly Ala Leu Thr Gly Ala Tyr Thr Ala Ala Ala Gln Gly Leu Thr
176 20 25 30
178 Thr Ile Val Leu Glu Lys Thr Asp Arg Phe Gly Gly Thr Ser Ala Tyr
179 35 40 45
181 Ser Gly Ala Ser Ile Trp Leu Pro Gly Thr Gln Val Gln Glu Arg Ala
182 50 55 60
184 Gly Leu Pro Asp Ser Thr Glu Asn Ala Arg Thr Tyr Leu Arg Ala Leu
185 65 70 75 80
187 Leu Gly Asp Ala Glu Ser Glu Arg Gln Asp Ala Tyr Val Glu Thr Ala
188 85 90 95
190 Pro Ala Val Val Ala Leu Leu Glu Gln Asn Pro Asn Ile Glu Phe Glu
191 100 105 110
193 Phe Arg Ala Phe Pro Asp Tyr Tyr Lys Ala Glu Gly Arg Met Asp Thr
194 115 120 125
196 Gly Arg Ser Ile Asn Pro Leu Asp Leu Asp Pro Ala Asp Ile Gly Asp

```

RAW SEQUENCE LISTING

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:31

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw

```

197      130      135      140
199 Leu Ala Gly Lys Val Arg Pro Glu Leu Asp Gln Asp Arg Thr Gly Gln
200 145      150      155      160
202 Asp His Ala Pro Gly Pro Met Ile Gly Gly Arg Ala Leu Ile Gly Arg
203      165      170      175
205 Leu Leu Ala Ala Val Gln Ser Thr Gly Lys Ala Glu Leu Arg Thr Glu
206      180      185      190
208 Ser Val Leu Thr Ser Leu Ile Val Glu Asp Gly Arg Val Val Gly Ala
209      195      200      205
211 Glu Val Glu Ser Gly Gly Glu Thr Gln Arg Ile Lys Ala Asn Arg Gly
212      210      215      220
214 Val Leu Met Ala Ala Gly Gly Ile Glu Gly Asn Ala Glu Met Arg Glu
215 225      230      235      240
217 Gln Ala Gly Thr Pro Gly Lys Ala Ile Trp Ser Met Gly Pro Phe Gly
218      245      250      255
220 Ala Asn Thr Gly Asp Ala Ile Ser Ala Gly Ile Ala Val Gly Gly Ala
221      260      265      270
223 Thr Ala Leu Leu Asp Gln Ala Trp Phe Cys Pro Gly Val Glu Gln Pro
224      275      280      285
226 Asp Gly Ser Ala Ala Phe Met Val Gly Val Arg Gly Gly Leu Val Val
227      290      295      300
229 Asp Ser Ala Gly Glu Arg Tyr Leu Asn Glu Ser Leu Pro Tyr Asp Gln
230 305      310      315      320
232 Phe Gly Arg Ala Met Asp Ala His Asp Asp Asn Gly Ser Ala Val Pro
233      325      330      335
235 Ser Phe Met Ile Phe Asp Ser Arg Glu Gly Gly Gly Leu Pro Ala Ile
236      340      345      350
238 Cys Ile Pro Asn Thr Ala Pro Ala Lys His Leu Glu Ala Gly Thr Trp
239      355      360      365
241 Val Gly Ala Asp Thr Leu Glu Glu Leu Ala Ala Lys Thr Gly Leu Pro
242      370      375      380
244 Ala Asp Ala Leu Arg Ser Thr Val Glu Lys Phe Asn Asp Ala Ala Lys
245 385      390      395      400
247 Leu Gly Val Asp Glu Phe His Arg Gly Glu Asp Pro Tyr Asp Ala
248      405      410      415
250 Phe Phe Cys Pro Pro Asn Gly Gly Ala Asn Ala Ala Leu Thr Ala Ile
251      420      425      430
253 Glu Asn Gly Pro Phe Tyr Ala Ala Arg Ile Val Leu Ser Asp Leu Gly
254      435      440      445
256 Thr Lys Gly Gly Leu Val Thr Asp Val Asn Gly Arg Val Leu Arg Ala
257      450      455      460
259 Asp Gly Ser Ala Ile Asp Gly Leu Tyr Ala Ala Gly Asn Thr Ser Ala
260 465      470      475      480
262 Ser Leu Ser Gly Arg Phe Tyr Pro Gly Pro Gly Val Pro Leu Gly Thr
263      485      490      495
265 Ala Met Val Phe Ser Tyr Arg Ala Ala Gln Asp Met Ala Lys
266      500      505      510
269 <210> SEQ ID NO: 3
270 <211> LENGTH: 158

```

RAW SEQUENCE LISTING

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:31

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw

```

271 <212> TYPE: DNA
272 <213> ORGANISM: Rhodococcus erythropolis
274 <400> SEQUENCE: 3
275 atcatcgatt atgtgtcccg gccgcgaacg accgcgctaa ttctctcacc tggaccaccc 60
277 atctcggcat attgccgct cagtgggacc tggcatggcc ttccagtgcc gtgcggtatt 120
279 ccgtggacac cccaccctct tggagtaagg acgcaatg 158
282 <210> SEQ ID NO: 4
283 <211> LENGTH: 19
284 <212> TYPE: DNA
285 <213> ORGANISM: Artificial Sequence
287 <220> FEATURE:
288 <223> OTHER INFORMATION: Description of Artificial Sequence:primer
290 <400> SEQUENCE: 4
291 ggcgacgttg ccgagaatt 19
294 <210> SEQ ID NO: 5
295 <211> LENGTH: 624
296 <212> TYPE: DNA
297 <213> ORGANISM: Rhodococcus erythropolis
299 <220> FEATURE:
300 <221> NAME/KEY: CDS
301 <222> LOCATION: (1)..(624)
303 <400> SEQUENCE: 5
304 atg ggg gcg acg ttg ccg aga att gcc gag gtc agg gac gct gct gag 48
305 Met Gly Ala Thr Leu Pro Arg Ile Ala Glu Val Arg Asp Ala Ala Glu
306 1 5 10 15
308 ccc agt tcg gac gag cag cgg gcg cgc cat gtg cgg atg ctg gaa gcg 96
309 Pro Ser Ser Asp Glu Gln Arg Ala Arg His Val Arg Met Leu Glu Ala
310 20 25 30
312 gcc gcc gaa ttg ggg acc gag aaa gaa ctc tca cgg gtt cag atg cac 144
313 Ala Ala Glu Leu Gly Thr Glu Lys Glu Leu Ser Arg Val Gln Met His
314 35 40 45
316 gaa gtt gcc aag cgg gca ggc gtg gcc atc ggc act ctc tac cgc tat 192
317 Glu Val Ala Lys Arg Ala Gly Val Ala Ile Gly Thr Leu Tyr Arg Tyr
318 50 55 60
320 ttc cct tcg aag acg cac ctc ttc gtc gct gtg atg gtc gag cag atc 240
321 Phe Pro Ser Lys Thr His Leu Phe Val Ala Val Met Val Glu Gln Ile
322 65 70 75 80
324 gat cag atc ggc gac agt ttc gcc aag cat cag gtg cag tcg gcc aat 288
325 Asp Gln Ile Gly Asp Ser Phe Ala Lys His Gln Val Gln Ser Ala Asn
326 85 90 95
328 ccg cag gac gcc gtg tac gag gtc ctg gtg cgc gcg act cgc ggg tta 336
329 Pro Gln Asp Ala Val Tyr Glu Val Leu Val Arg Ala Thr Arg Gly Leu
330 100 105 110
332 ctg cgt cgg ccg gcc ctt tcg act gcg atg ctg cag tcg tcc agt acc 384
333 Leu Arg Arg Pro Ala Leu Ser Thr Ala Met Leu Gln Ser Ser Ser Thr
334 115 120 125
336 gcc aac gtc gcg acg gtg ccg gac gtg ggc aag atc gat cgc ggc ttc 432
337 Ala Asn Val Ala Thr Val Pro Asp Val Gly Lys Ile Asp Arg Gly Phe
338 130 135 140

```

VERIFICATION SUMMARY

DATE: 03/06/2006

PATENT APPLICATION: US/10/537,201A

TIME: 15:23:32

Input Set : A:\Seq. Listing. 2002-744US.txt

Output Set: N:\CRF4\03062006\J537201A.raw